

Cloud Computing

Virtual Resources and Distributed Cloud Applications

Chapter 1: Introduction



Distributed and Operating Systems
Electrical Engineering and Computer Science
Technische Universität Berlin

Who are we?

- Prof. Dr. Odej Kao
 - Lectures
 - Odej.kao@tu-berlin.de
- Dominik Scheinert
 - Practical assignments, Q&A
 - dominik.scheinert@tu-berlin.de



Course Organization

	When	Where
Lectures	Weekly Tuesday 14:00h	MA001
Assignments Q&A	Every 3rd week Tuesday 16:00h	H 2013

- Lectures
 - PDFs with the slides published prior to the start of the chapter on ISIS
 - Videos from previous semesters available (although some lecture contents changed)
- Practical Assignments
 - Assignments released every three weeks on Mondays
 - Four practical assignments in total
 - Q&A, like tutorials, mainly for questions and discussion -> see ISIS schedule

Examination

- Portfolio exam
 - 50% written exam at the end
 - 50% for the four exercises, each weighted equally
 - Exercise points from previous attempts may be used (write to Dominik Scheinert before the first exercise is due)
 - Grading using “Notenschlüssel 1”
- Exercise assignments will cover practical application of infrastructure and platform technologies
- Written exam: 22.02.2024 (1. appointment), 12.04.2024 (2. appointment)

Practical Assignments

- Five tutorial sessions (every ~3th Tuesday, 4-6pm)
 - 07.11.2023 (Assignment 1)
 - 28.11.2023 (Assignment 2)
 - 19.12.2023 (Assignment 3)
 - 09.01.2024 (Assignment 3)
 - 23.01.2024 (Assignment 4)
- Each session discusses one assignment
- Each assignment is due ~1,5 weeks later (skipping christmas)
- Assignments are solved in teams of 5 students
- Assignment material is published on ISIS

Practical Assignments

- Topics covered
 - Usage of Infrastructure-as-a-service clouds (dashboard and CLI API)
 - Benchmarking the performance of different platforms
 - Benchmarking the cloud API
 - DevOps tasks for cloud services
 - Infrastructure-as-code orchestration
 - Container virtualization (Docker)
 - Container-based orchestration (Kubernetes)
- Sources of Information
 - Assignment sheets
 - Online documentations
 - Online forums (e.g. stackoverflow)
 - ISIS forum: ask your fellow students!

Practical Assignments

What to do next (before the first assignment!):

- Find a group, organize on ISIS (deadline: 29.10.2023)
- Create account on Google Cloud Platform
 - We will provide each one of you with a 50\$ coupon

What to keep in mind:

- Register on MOSES (deadline: 17.11.2023)

Lecture Resources

- ISIS course for announcements, lecture recordings, slides, discussions, and exercise groups:
- <https://isis.tu-berlin.de/course/view.php?id=35494>
- Please help yourself and each others!
- Recommended reading:
 - Tanenbaum: Modern Operating Systems
 - Humble, Farley: Continuous Delivery
 - Kleppmann: Designing Data-Intensive Applications
 - ... more literature announced during the course!

Module “Cloud Computing”

- Credits (according to ECTS): 6 points
- Limited to 120 participants
- Accepted participants can register for the exam on MOSES and the examination office

Background

Clusters

- Many locally connected computers

Grids

- Loosely coupled and widely distributed computers or clusters

Clouds

- IT resources as a utility

Fog

- Cloud services in closer proximity to users and edge devices

Clusters

- Mostly homogeneous (commodity) compute resources and software stacks
- Interconnected by a low-latency and high-bandwidth (local area) network
- Goal: improving availability, resource utilization, price/performance
- Examples
 - Analytics cluster at Facebook and Yahoo!

Grids

- Widely distributed heterogeneous compute resources
- Connected via a slow network (i.e. the internet)
- Heterogeneous software stacks unified by a middleware
- Examples
 - Worldwide LHC Computing Grid
 - BOINC

NIST Definition of Cloud Computing

- Definition of Cloud Computing according to NIST [1]:

“Cloud computing is a model for enabling **ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources** (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

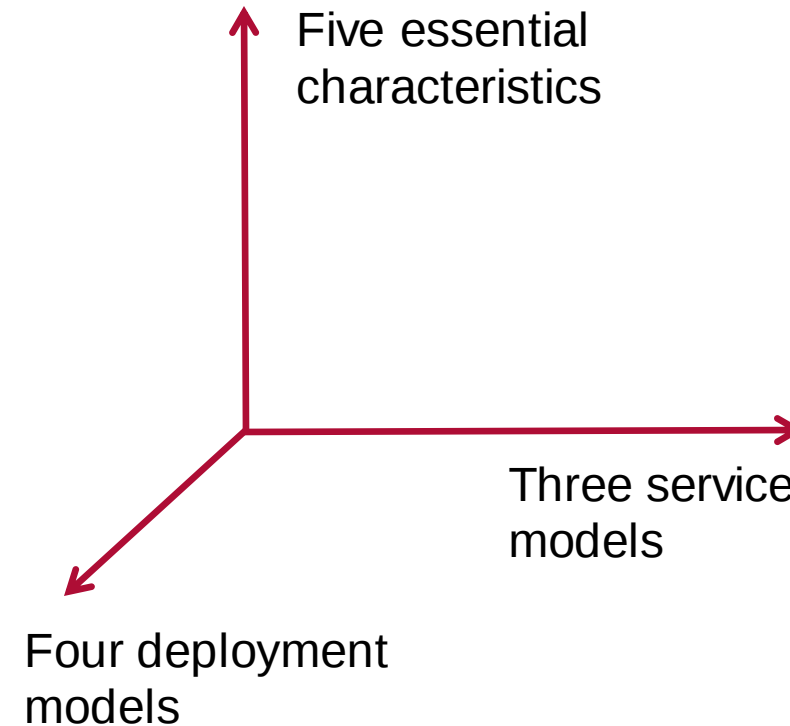
This cloud model is composed of five essential characteristics, three service models, and four deployment models.”

Standard 800-145

- NIST: National Institute of Standards and Technology
 - Agency of U.S. Department of Commerce
 - Responsible for standardization processes

Dimensions of Cloud Computing (NIST)

- NIST: “Cloud model is composed of
 - Five essential characteristics
 - Three service models
 - Four deployment models”



Five Characteristics of Cloud Computing (NIST)

- On-demand self-service
 - No human interaction required for resource provisioning
- Broad network access
 - Accessible over network with standard mechanisms
- Resource pooling
 - Pooled resources dynamically shared among several consumers, location independence
- Rapid elasticity
 - Capabilities can be provisioned/released on demand
- Measured service
 - Resource usage is monitored, controlled, and reported

Three Service Models of Cloud Computing (NIST)

- Software as a Service (SaaS)
 - Provider's application runs on cloud infrastructure, accessed over the network
 - Consumer does not control/manage underlying infrastructure
- Platform as a Service (PaaS)
 - Consumer deploys custom applications onto cloud infrastructure using programming languages, libraries, services and tools supported by provider
 - Consumer does not control/manage underlying infrastructure
- Infrastructure as a Service (IaaS)
 - Provider provisions processing, storage, network resources to consumer
 - Consumer does not control/manage underlying infrastructure but has control over operating systems, storage, and deployed applications

Four Deployment Models of Cloud Computing (NIST)

	Private Cloud	Community Cloud	Public Cloud	Hybrid Cloud
User of the cloud infrastructure?	Single organization	Organizations with shared concerns	Open for the general public	Composition of private / community / public cloud: - Clouds remain distinct, but bound together by standard mechanisms - Goal: Enable data / application portability
Owner of the cloud infrastructure?	Organization, third party, combination thereof	Organizations, third party, combination thereof	Business, academic, government organization, combination thereof	
Location of the cloud infrastructure?	On premise, off premise	On premise, off premise	On premise of cloud provider	

Examples

- AWS EC2
 - DigitalOcean Droplets
 - Google Kubernetes Engine
 - AWS S3
 - Amazon EMR
 - AWS Lambda
 - Azure Machine Learning Studio
 - AWS SageMaker
 - IBM Watson
-
- Google Mail, Google Docs, MS Office 365...

Background

Clusters

- Many locally connected computers

Grids

- Loosely coupled and widely distributed computers or clusters

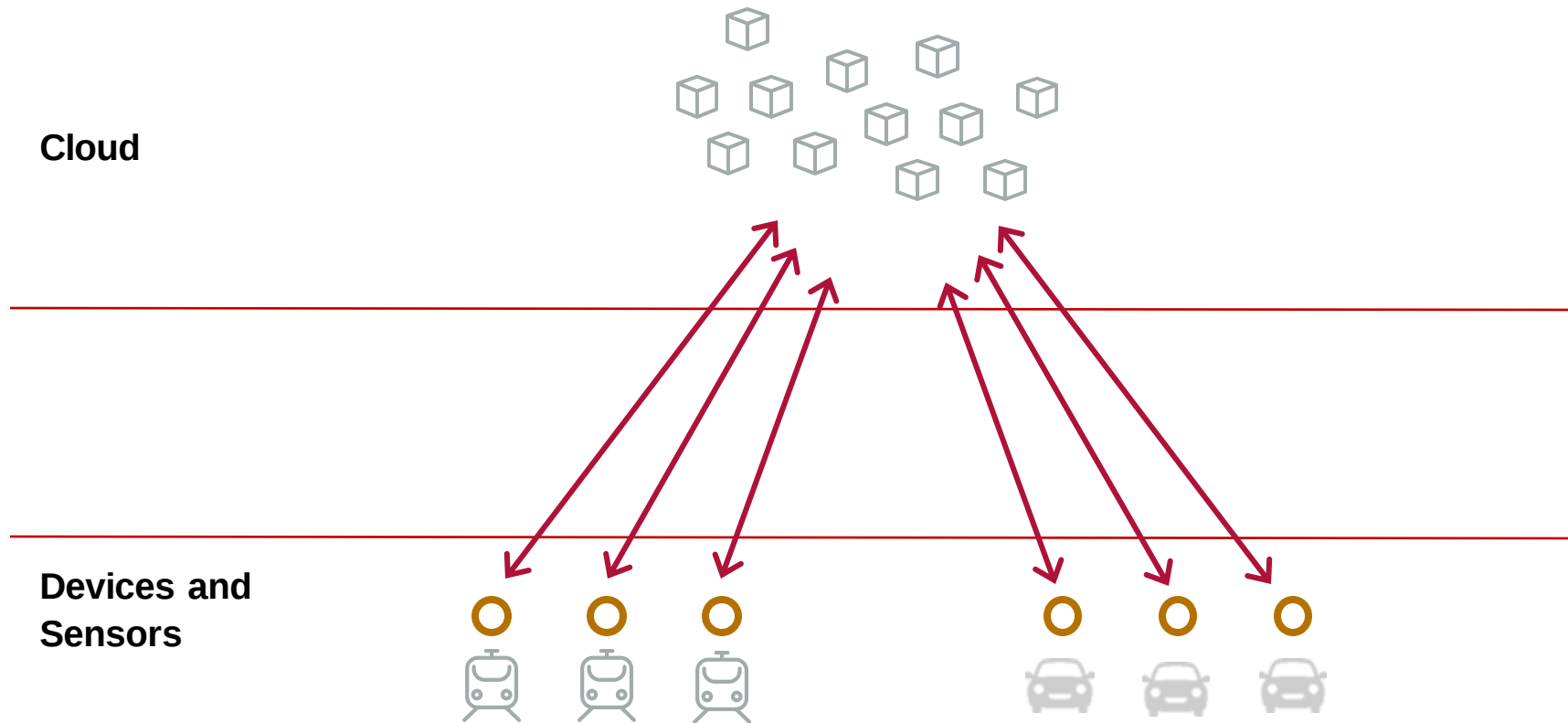
Clouds

- IT resources as a utility

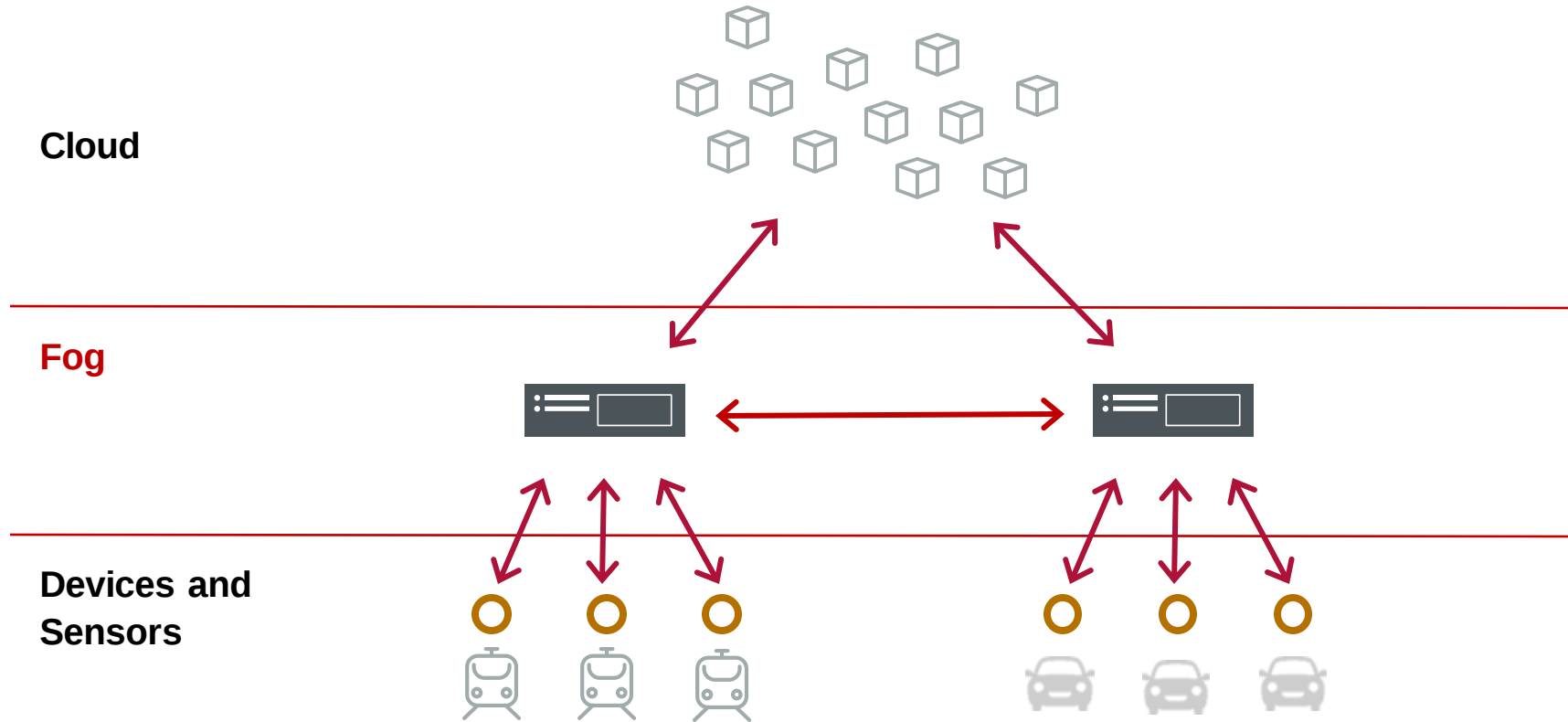
Fog

- Cloud services in closer proximity to users and edge devices

Cloud and the IoT



Cloud and the IoT: Fog Computing



NIST Definition of Fog Computing [2]

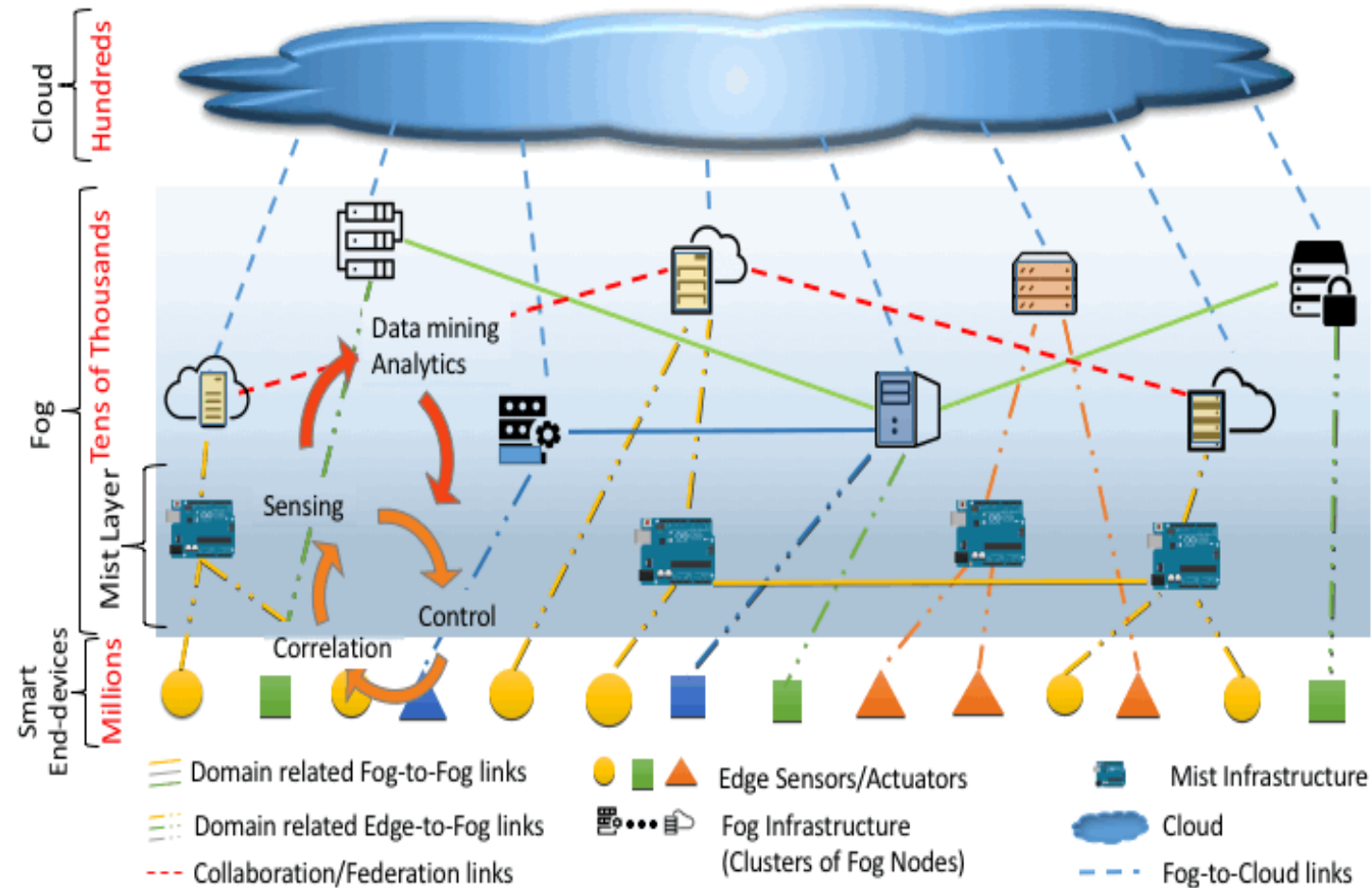
“Fog computing is a layered model for enabling **ubiquitous access to a shared continuum of scalable computing resources.**

The model facilitates the **deployment of distributed, latency-aware applications and services**, and consists of **fog nodes (physical or virtual)**, residing **between smart end-devices and centralized (cloud) services**. The fog nodes are context aware and support a **common data management and communication system.**

Fog computing **minimizes the request-response time** from/to supported applications, and provides, for the end-devices, **local computing resources** and, when needed, **network connectivity to centralized services.**

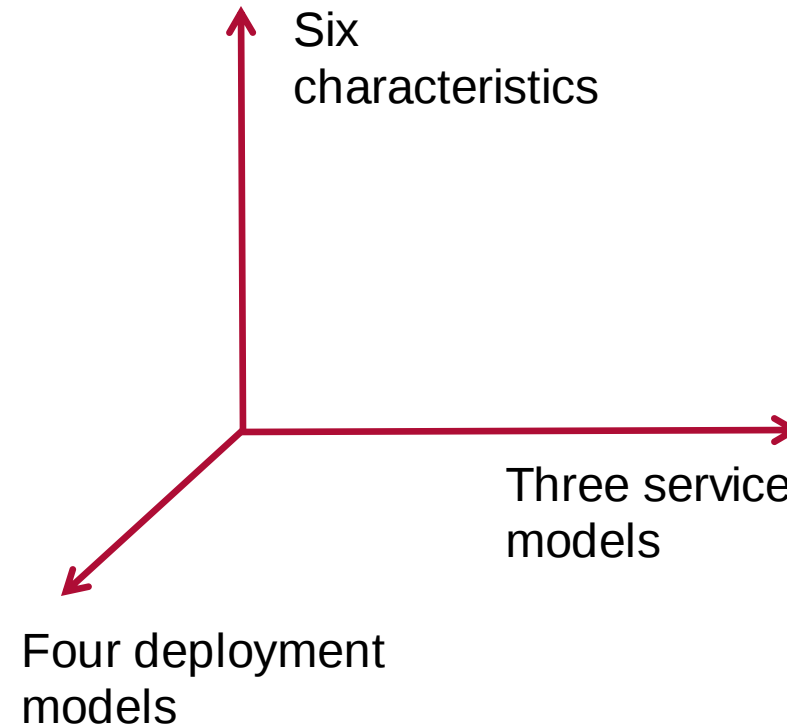
Standard 500-325

NIST Definition of Fog Computing (Ctnd.)



Dimensions of Fog Computing (NIST)

- Architectural service types:
 - SaaS
 - PaaS
 - IaaS
- Node deployment models:
 - private nodes
 - community nodes
 - public nodes
 - hybrid nodes



Characteristics of Fog Computing (NIST)

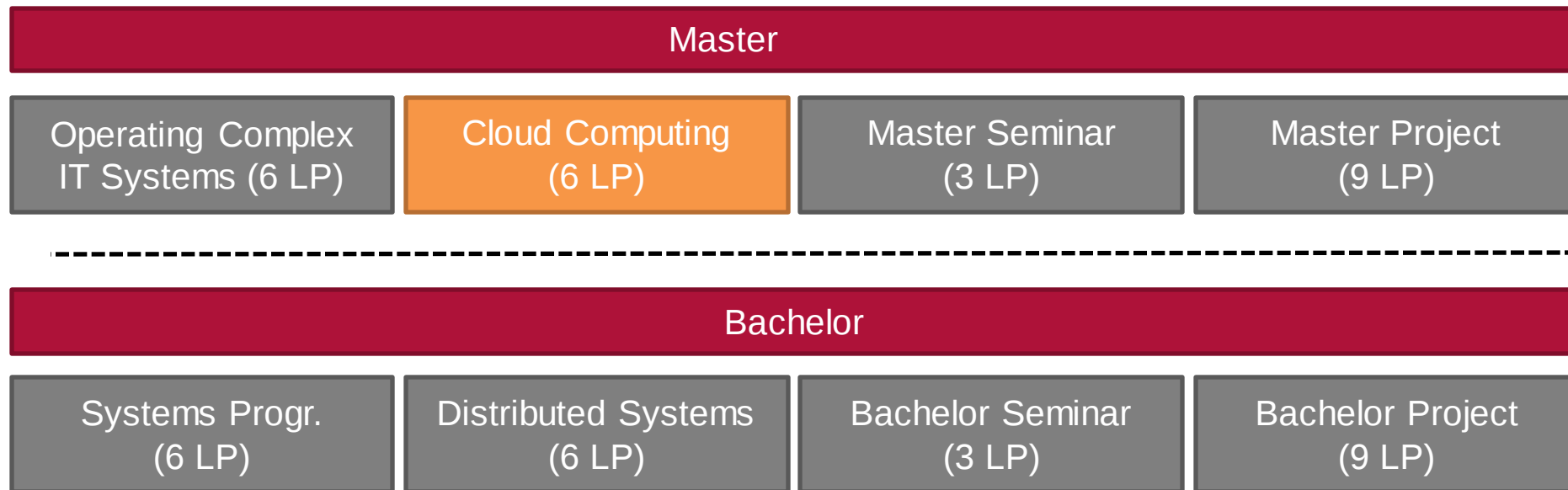
- Contextual location awareness, and low latency
- Geographical distribution
- Heterogeneity
- Interoperability and federation
- Real-time interactions
- Scalability and agility of federated, fog-node clusters

What the Lecture will Discuss

- Goal: Understand what Cloud Computing entails on a technical level
 - Know the different levels of abstractions
 - Understand the implications of resource sharing
 - Learn to take advantage of Cloud platforms
- Content
 2. Virtual Resources
 3. Management of Virtual Resources
 4. Scalable and Fault-Tolerant Cloud Applications
 5. Data-Intensive Cloud Applications
 6. Platforms and Serverless Computing
 7. Outlook Beyond Cloud

Teaching at DOS

- DOS curriculum
 - Bachelor: Basics of distributed and operating systems
 - Master: Advanced topics and stronger research focus



References

- [1] P. Mell, T. Grance: “The NIST Definition of Cloud Computing”, Technical Report, National Institute of Standards and Technology, 2011, <https://csrc.nist.gov/publications/detail/sp/800-145/final>
- [2] M. Iorga, L. Feldman, L., R. Barton, M. J. Martin, N. Goren, C. Mahmoudi: "Fog Computing Conceptual Model", Technical Report (Draft), National Institute of Standards and Technology, 2018, <https://www.nist.gov/publications/fog-computing-conceptual-model>